

Metode-metode *data mining* untuk penyelesaian masalah kehamilan dan persalinan

Annas Gading Pertiwi¹, Utomo Pujiyanto²

Jurusan Teknik Elektro, Universitas Negeri Malang, Malang, Indonesia

¹ annasgading@gmail.com; ² utomo.pujiyanto.ft@um.ac.id

INFORMASI ARTIKEL

Histori Artikel

Diterima : 22 Januari 2020
Direvisi : 19 Februari 2020
Diterbitkan : 4 April 2020

Kata Kunci:
Data Mining
Kehamilan
Persalinan

ABSTRAK

Keakuratan data yang digunakan pada bidang kesehatan dituntut untuk memiliki tingkat akurasi yang tinggi. Oleh karena itu banyaknya data pada bidang kesehatan ini perlu diimbangi dengan pemrosesan data yang sesuai salah satunya dengan menggunakan data mining. Data mining dapat digunakan untuk menggali informasi-informasi dari banyaknya data yang telah ada. Pada penelitian terdahulu khususnya pada masalah kehamilan dan persalinan, beberapa metode data mining telah diaplikasikan. Dalam data mining, terdapat metode prediksi yang juga kerap digunakan untuk menggali data di bidang kehamilan dan persalinan. Pada paper ini, akan dijelaskan kelebihan dan kelemahan metode-metode prediksi yang sering digunakan pada masalah kehamilan dan persalinan.

2019 SAKTI – Sains, Aplikasi, Komputasi dan Teknologi Informasi.

Hak Cipta.

I. Pendahuluan

Data mining yang telah dikembangkan sejak tahun 1990 mempunyai potensial yang besar untuk menemukan informasi-informasi atau pola dari informasi yang belum diketahui. Ini dikarenakan banyaknya data akurat yang tercatat pada bidang kesehatan tetapi data-data tersebut belum terlalu diolah oleh *expert*. Data mining dapat digunakan menggali informasi dari data-data yang belum diolah tersebut. Metode data mining dapat diklasifikasikan berdasarkan cara pengerjaannya yaitu *supervised learning*, *unsupervised learning* dan *semi-supervised learning*. Data mining pada umumnya dapat diklasifikasikan sebagai deskripsi dan prediksi. Proses deskripsi bertujuan untuk menemukan pola yang dapat dipahami manusia dan asosiasi, sedangkan metode prediksi bertujuan untuk meramalkan suatu hal [1].

Kehamilan adalah kondisi ketika seorang wanita mengandung embrio didalam rahimnya. Pada masa kehamilan, ibu dan anak menghadapi banyak resiko kesehatan karena pada masa kehamilan, sang ibu menjadi lebih sensitif dan rapuh. Komplikasi dari kehamilan dan proses mengandung merupakan penyebab kedua dari kematian pada gadis berumur 15-19 di dunia [2]. Pada tahun 2015, diperkirakan sebanyak 303 ribu wanita meninggal dikarenakan permasalahan terkait dengan kehamilan dan 2,7 juta bayi meninggal pada usia 28 hari pertama dan 2,6 juta bayi meninggal *stillbirth*, yaitu meninggal di janin atau selama persalinan [3].

Data mining telah banyak diaplikasikan pada bidang kesehatan, termasuk pada masalah kehamilan dan persalinan. Pada bidang kesehatan, banyak dilakukan metode *Supervised Learning*, karena kebanyakan data dari kesehatan telah dilengkapi dengan label, dan dibutuhkan data training untuk menentukan hasil dari proses tersebut. Salah satu teknik data mining yang banyak digunakan dalam bidang ini adalah teknik prediksi. Salah satu metode yang terkenal adalah *Naïve Bayesian*. Pada pengklasifikasian penyakit kandungan, *Naïve Bayesian Classifier* dengan *Laplacian Smoothing* memiliki tingkat akurasi lebih tinggi daripada *Learning Vector Quantization*. [4] *Naïve Bayesian* telah terbukti dapat bekerja dengan sangat baik pada bidang kesehatan [5], [6]. Karena induksi datanya cepat, metode ini kerap digunakan sebagai algoritma pendasar dalam studi komparatif [1]. *Naïve Bayesian* dikenal lebih mudah dipahami daripada algoritma lain, seperti *Neural Networks* dan *Support Vector Machine* [7].

Decision tree adalah metode data mining yang digambarkan dengan *nodes*, *branches*, dan *leaves*, dimana penentuan kedudukan *nodes* dipengaruhi oleh hasil *test* yang dilakukan. *Decision tree* dapat digunakan untuk metode klasifikasi [8], [9] dan prediksi [1], [10], [11]. *Decision tree* juga telah terbukti bekerja dengan baik pada bidang kesehatan [1], [12]. *Decision tree* mudah ditafsirkan, bahkan oleh masyarakat awam dan mudah untuk diikuti. Metode ini dapat mengatasi nilai yang hilang dan bisa menggabungkan data heterogen menjadi sebuah model [12]. Selain metode-metode tersebut, terdapat pula metode *Support Vector Machine* [13], *Artificial*

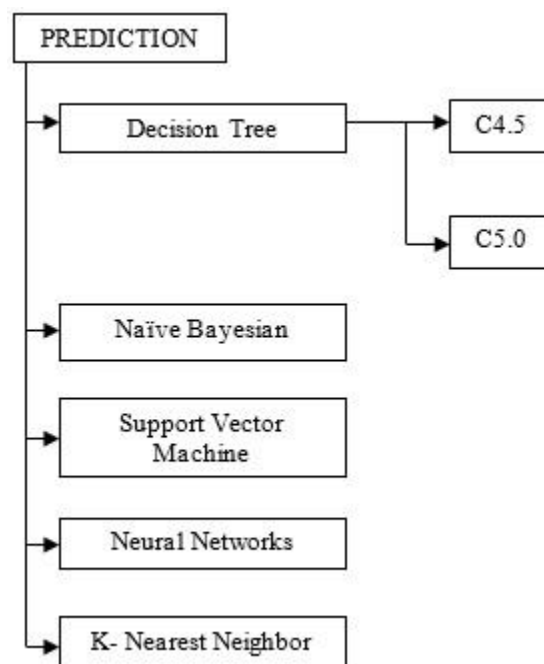
Neural Network [14], [15] dan *K-Nearest Neighbor* [16] yang juga sering digunakan oleh peneliti pada bidang ini.

Tujuan dari paper ini adalah untuk mengkaji ulang kondisi seni pada metode prediksi *data mining* yang digunakan pada masalah kehamilan dan persalinan serta untuk menampilkan kelebihan dan kekurangan dari berbagai algoritma untuk dijadikan perbandingan. Paper ini berisi tentang peninjauan teknik prediksi, perbandingan antar metode, dan menyimpulkan algoritma prediksi yang efisien untuk masalah kehamilan dan persalinan.

II. Tinjauan Teknik-Teknik Prediksi

Pada bidang kesehatan, data mining digunakan untuk analisis dan prediksi dari bermacam-macam penyakit [1], [13], [17]–[20]. Secara luas, metode data mining dapat diklasifikasikan berdasarkan apa yang dihasilkan seperti, deskripsi dan visualisasi; asosiasi dan klustering; dan klasifikasi dan estimasi, yang termasuk dalam model prediktif [21]. Data mining dibagi menjadi dua kategori, yaitu kategori prediksi yang terdiri dari klasifikasi, regresi dan *times series* dan kategori *knowledge discovery* yang terdiri dari deteksi deviasi, klustering, aturan asosiasi, penyimpulan, visualisasi dan *text mining*[5]. Metode prediksi adalah metode yang digunakan untuk memprediksi label apa yang akan diberikan pada data. Metode prediksi data mining bekerja sama seperti seorang manusia menganalisa data, perbedaannya adalah Metode prediksi data mining dapat digunakan pada sebuah dataset yang banyak, namun manusia hanya dapat menganalisa dataset yang kecil [22], [23]. Metode ini belajar dari *data training* untuk menemukan informasi, kemudian mengaplikasikannya pada *data testing* sehingga disebut ‘prediksi’.

Langkah-langkah yang dilakukan dalam metode prediksi data mining terdiri dari tiga tahap, yaitu *data processing*, *prediction* dan *deployment*. Pada tahap *data processing* mengumpulkan data, kemudian dilakukan pemeriksaan preliminary pada kualitas data, kemudian dilakukan pemeriksaan apakah ada hal-hal seperti *relationships*, korelasi, atau interaksi kuat antar atribut. Jika ada, maka dilakukan train set, pembagian data dan tes validasi. Jika tidak ada, maka dilakukan pengukuran kembali. Kemudian, pada tahap *prediction*, dilakukan metode prediksi, kemudian ditentukan apakah diperlukan pengukuran kembali, jika iya, maka kembali ke tahap selanjutnya. Apabila tidak, validasi model, kemudian membandingkan model sehingga menghasilkan model final. Selanjutnya pada tahap *deployment*, model final tersebut diaplikasikan ke masalah yang ada [23]. Seperti yang telah disebutkan, terdapat banyak metode prediksi yang diajukan pada literatur ini sebagai pada Gambar 1. Metode-metode tersebut dipakai untuk memecahkan masalah kehamilan dan persalinan.



Gambar. 1. Algoritma Prediksi yang Diajukan

A. Decision Tree

Decision tree adalah sebuah model yang direpresentasikan sebagai sebuah pohon, dimana *nodes*-nya merupakan atribut dari data dan *rules*-nya dituliskan pada cabang-cabangnya, sehingga terlihat seperti *flowchart*. Sebuah *binary decision tree* memisahkan data (*parent node*) menjadi dua subset (*child nodes*) dengan menghitung *feature* terbaik yang ditentukan oleh sebuah kriteria yang dipilih. Dua subset yang terpilih itu menjadi *parent nodes* yang baru dan dibagi lagi menjadi dua *child nodes*. Pemecahan ini akan dilanjutkan sampai

semua observasi terklasifikasikan [12]. Algoritma yang digunakan untuk *decision tree* pada masalah kehamilan dan persalinan adalah C4.5 [8], [9], [24], [25] dan C5.0 [9][14]. Berikut adalah algoritma dasar untuk menginduksi sebuah pohon keputusan dari data training [26].

```

1.  Buat sebuah node N;
2.  IF data di D sama dengan pada C, THEN
3.    RETURN N sebagai leaf node yang dilabeli kelas C;
4.  IF attribute_list kosong THEN
5.    RETURN N sebagai leaf node yang dilabeli kelas D //majority voting
6.  APPLY Attribute_selection_method(D, attribute_list)
7.  Labeli node N dengan splitting_criterion;
8.  IF splitting_attribute adalah nilai diskrit AND
   multiway split diperbolehkan THEN
9.    attribute_list <- attribute_list – splitting_attribute
10. FOR EACH output j dari splitting_criterion
11.   jadikan Dj menjadi data di D outcome;
12.   IF D, kosong, maka
13.     attach sebuah leaf yang dilabeli majority class di D ke node N;
14.   ELSE attach node yang dihasilkan Generate_decision_tree ke node N;
   ENDFOR.
RETURN N;
```

Selanjutnya adalah pemodelan dengan *decision tree*. Berikut adalah permodelan *decision tree* dengan menggunakan berbagai macam algoritma.

1) C4.5

Algoritma C4.5 dikenalkan pada tahun 1993 oleh Ross Quinlan, yang juga mengenalkan algoritma ID3 pada kita [27]. Algoritma C4.5 ini merupakan pengembangan dari algoritma ID# yang sangat sensitive pada fitur dengan jumlah nilai yang banyak. Algoritma C4.5 menggunakan “*Information gain*”, yang direpresentasikan pada persamaan (1) dan (2).

$$Gain(p) = F(Info(T) - Info(p.T)) \quad (1)$$

Dimana:

$$Info(p.T) = \sum_{j=1}^n (p_j \times Entropie(p_j))$$

$$Info(T) = Entropy(T)$$

F = jumlah sampel pada *database* dengan nilai yang diketahui atau total dari sampel pada sebuah set.

Decision trees dibangun oleh algoritma C4.5 dengan menggunakan *data training* dengan memilih satu atribut yang sangat efektif membagi sampel-sampel menjadi subset yang diperkaya pada satu kelas atau lainnya. Atribut dengan nilai *information gain* tertinggi dipilih untuk membuat keputusan. Algoritma C4.5 melakukan pengerjaan yang sama seperti ID3, namun dapat diimprovisasi menjadi empat pekerjaan yaitu (a) apat menggunakan data yang continuous, (b) dapat menggunakan missing values, (c) mampu menggunakan atribut dengan weight yang berbeda, dan pruning tree saat setelah dibuat [27].

2) C5.0

C5.0 atau See5 merupakan pengembangan dari C4.5. C4.5 merupakan algoritma yang mengikuti aturan dari ID3, dan C5 juga mengikut aturan dari C4.5, sehingga mempunyai *rules* yang hampir sama dengan algoritma C4.5. Sama seperti algoritma C4.5, algoritma C.5 menyediakan seleksi fitur, *cross validation*, dan pengurangan *error pruning* [28]. C5.0 juga mempunyai empat fitur, diantaranya (a) decision tree yang besar dapat dilihat sebagai aturan-aturan yang mudah dipahami, (b) memberikan pengetahuan tentang noise dan data yang hilang, (c) permasalahan over fitting dan error pruning diselesaikan oleh algoritma C5.0, dan (d) pengklasifikasian C5.0 dapat mengantisipasi atribut mana yang relevan dan tidak [29].

B. Naïve Bayesian

Metode *Naïve Bayesian* terkenal dengan prediksi probabilitas dari suatu data yang disajikan. Seperti namanya, metode *Bayesian* berdasarkan teorema *Bayes*. Sedangkan metode *naïve Bayes* juga sering disebut *idiot's Bayes*, *simple Bayes* dan *independence Bayes* karena sangat mudah untuk dibuat dan tidak membutuhkan proses iterasi yang kompleks [30]. *Bayesian Network* digunakan oleh banyak peneliti pada bidang kesehatan [31]. *Naïve Bayes* menggunakan semua atribut, yang berdasarkan pada kedua asumsi, yaitu (a) Semua atribut sama penting [28], dan (b) semua atribut bersifat independen, tidak berhubungan antar satu dengan sama lain [28], [31].

Asumsi independensi yang kuat antar parameter membuat syarat peluang menjadi sangat sederhana dan mudah untuk dihitung. Persamaan 3 merupakan model dari teorema *Naïve Bayes*. Teorema *Naïve Bayes* dapat digunakan sebagai metode untuk klasifikasi.

$$P(F_1 \dots F_n | C) = \prod_{i=1}^n P(F_i | C) \quad (3)$$

C. Support Vector Machine

Support Vector Machine (SVM) adalah metode untuk klasifikasi untuk data *linear* dan *nonlinear*. Metode SVM bekerja dengan pemetaan *nonlinear* untuk mengubah *training* menjadi dimensi yang lebih tinggi, kemudian dicari *linear optimal separating hyperplane*. Selanjutnya, data dari dua kelas tersebut dapat dipisahkan oleh sebuah *hyperplane*, yang dicari menggunakan *support vectors* dan *margins* [26]. *Support vector machines* diprediksikan adalah algoritma paling baik dalam akurasi [32]. *Support vector machines* dinyatakan sebagai algoritma yang kuat karena strukturnya yang sederhana dan memerlukan jumlah fitur yang lebih sedikit.

D. Neural Network

Artificial neural network merupakan metode kecerdasan buatan berbasis data *modelling* yang populer dan banyak digunakan pada bidang kesehatan [1]. Ini dikarenakan kinerja prediktifnya, meskipun juga mempunyai beberapa kekurangan, diantaranya sensitivitasnya yang tinggi pada parameter dari metodenya [33]. *Neural network* (NN) dapat digunakan untuk klasifikasi dan pengenalan pola. Sebuah NN dapat beradaptasi karena dapat mengubah strukturnya dan mengatur beratnya berdasarkan informasi dari jaringan pada fase *learning* untuk meminimalisir *errors* [31].

E. K-Nearest Neighbor

Metode klasifikasi *K-Nearest Neighbour* (K-NN) merupakan salah satu metode dari algoritma yang paling sederhana yang menemukan data yang tidak teridentifikasi menggunakan *data points* yang diketahui (*nearest neighbor*) dan mengklasifikasikan data dengan sistem voting [31]. Jika ditambahkan sebuah data yang tidak diketahui labelnya, K-NN mengklasifikasikan data tersebut dengan menghitung jarak terdekat *majority* dengan data yang sudah ada. Untuk menghitung jarak terdekat tersebut, digunakan banyak metode, namun yang paling banyak digunakan adalah *Euclidean distance*.

Karena sangat sederhana, metode ini telah banyak digunakan pada berbagai bidang, seperti *Pattern recognition*, *Image databases*, *Internet marketing*, *Cluster analysis* dan lainnya [34]. Agar tidak terjadi *voting* seri pada *binary classification*, maka diusahakan nilai *k* berupa bilangan ganjil. Berikut adalah *pseudocode* dari metode *K-Nearest Neighbor*.

```
K <- jumlah dari nearest neighbors
foreach object X in test set do
    hitung jarak D(X,Y) antara X dan setiap object Y
    neighborhood <- k neighbors pada data training yang
        terdekat dengan X
    X class <- SelectClass(neighborhood)
Endfor
```

III. Perbandingan Metode-Metode yang Ada

Pada sesi ini, dipaparkan kelebihan dan kekurangan dari berbagai macam algoritma prediktif yang telah dijelaskan diatas. Tabel 1 menunjukkan perbandingan dari beberapa metode yang telah dijabarkan pada sesi sebelumnya. Beberapa metode tersebut diantaranya Decision Tree C4.5, Decision Tree C5.0, *Naïve Bayesian*, *Support Vector Machine*, *Neural Network*, dan *K- Nearest Neighbor*.

Tabel 1 Perbandingan Metode-Metode

Metode	Kelebihan	Kelemahan
<i>Decision Tree</i> C4.5	<p>Efektif pada penyimpanan memori, karena memerlukan memori lebih sedikit untuk mengeksekusi program yang besar.</p> <p>Relatif cepat.</p> <p>Dapat mengatasi <i>missing attributes</i>.</p> <p>Dapat menggunakan data yang continuous [27].</p> <p>Dapat menentukan seberapa dalam untuk membuat <i>decision tree</i> [28].</p> <p>Transparan dalam menghasilkan nilai [1].</p>	<p>C4.5 membuat <i>empty branches</i> [35].</p> <p>Permasalahan <i>overfitting</i>.</p> <p>Rentan terhadap <i>noise</i> atau <i>outliers</i> [35].</p>

Metode	Kelebihan	Kelemahan
<i>Decision Tree</i> C5.0	Lebih cepat dan efektif pada penyimpanan memori daripada C4.5. Dengan <i>decision tree</i> yang kecil, dapat dihasilkan hasil yang baik [29]. Pada dasarnya merupakan pengembangan dari algoritma C4.5, sehingga pada beberapa aspek lebih unggul daripada algoritma C4.5	Terdapat masalah segmentasi yang biasanya karena algoritma C5.0 menghabiskan <i>stack space</i> yang telah disediakan [36]. <i>Weighting case</i> tidak menjamin hasil menjadi lebih akurat [36].
<i>Naïve Bayesian</i>	Metodenya sederhana sehingga mudah dalam pengimplementasiannya. <i>Performance</i> -nya baik. [34] Akurat dan berakurasi tinggi bila diimplementasikan pada data yang banyak. [4]	Menganggap semua atribut independen, sedangkan pada kesehatan, terkadang terdapat atribut yang saling berhubungan. Hasil kurang akurat jika <i>data training</i> sangat kurang dibandingkan <i>data test</i> . [4]
<i>Support Vector Machine</i>	Efektif pada <i>high dimensional space</i> . Dapat menangani <i>non-linear</i> data. Dapat menangani permasalahan <i>multi class</i> [28]. Memberikan solusi yang unik karena optimitas masalahnya <i>convex</i> [37]. Tidak adanya asumsi tentang bentuk fungsional dari transformasi [37]. Tidak terlalu bergantung pada seleksi parameter yang spesifik [1].	Komputasinya mahal [34], karena waktu <i>training data</i> relative lama [38]. Jika <i>features</i> sangat lebih banyak dari <i>samples</i> , kinerja akan menjadi buruk. Pemilihan <i>parameters</i> [38].
<i>Neural Network</i>	Memerlukan <i>formal stastical training</i> lebih sedikit dibanding algoritma lainnya. Mampu mendeteksi relasi nonlinear yang kompleks antara variabel <i>dependent</i> dan <i>independent</i> [39]. Dapat mendeteksi semua interaksi yang memungkinkan antara variabel predictor [39]. Mampu melakukan <i>multiple training</i> [39].	Kebiasaan “black box” [39]. Proses komputasinya lebih besar [39]. Rawan terjadi overfitting [39].
<i>K- Nearest Neighbor</i>	Mudah dipahami dan diimplementasikan. Proses <i>data training</i> dilakukan dengan cepat. Tidak terpengaruh oleh <i>noise</i> [34]. Cocok untuk kelas <i>multimodal</i> [34].	Memori terbatas, tidak dapat memproses data yang terlalu banyak. Sensitif pada <i>local structure</i> dari data [34]. Termasuk salah satu <i>lazy learners</i> .

IV. Kesimpulan

Pada bidang kehamilan dan persalinan, terdapat banyak data yang belum terolah oleh peneliti. *Data mining* merupakan metode yang cocok untuk mengolah data dalam jumlah besar sehingga dapat diambil informasinya. Metode *data mining* pada bidang kehamilan dan persalinan yang lebih sering digunakan adalah metode *supervised learning*, karena perlu adanya pengecekan dikarenakan ini berhubungan dengan kesehatan. Salah satu metode dari *data mining*, metode prediksi, telah dikaji di paper ini beserta algoritma-algoritma yang sering digunakan untuk penelitian.

Dari kajian pada paper ini, dapat disimpulkan alasan dipilihnya algoritma C4.5 sebagai algoritma yang cocok untuk mengolah data kehamilan dan persalinan. Meskipun algoritma C4.5 mempunyai kelemahan yaitu rentan terhadap *noise* atau *outliers*, tetap dapat digunakan karena pada data kesehatan, tidak terlalu banyak *outliers*, selain itu terdapat berbagai cara untuk mengatasi kelemahan tersebut. Kelemahan lain adalah algoritma C4.5 cenderung membuat *branches* yang tidak terlalu penting, ini juga tidak terlalu menjadi masalah pada bidang kehamilan dan persalinan karena atribut yang dipilih merupakan atribut yang penting.

Daftar Pustaka

- [1] R. Bellazzi and B. Zupan, “Predictive data mining in clinical medicine: Current issues and guidelines,” *Int. J. Med. Inform.*, vol. 77, no. 2, pp. 81–97, 2008.
- [2] World Health Organization, “WHO | Adolescent pregnancy,” *WHO*, World Health Organization, 2016.
- [3] World Health Organization, “WHO | Pregnant women must be able to access the right care at the right time, says WHO,” *WHO*, World Health Organization, 2016.
- [4] P. A. Nugraha, R. Saptono, and M. E. Sulisty, “Perbandingan Metode Probabilistik Naïve Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan,” vol. 2, no. 2, pp. 20–33, 2013.
- [5] I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. 2011.
- [6] E. H. Shortliffe and J. J. Cimino, *Biomedical informatics: computer applications in health care and biomedicine*. .
- [7] M. a. Hall and G. Holmes, “Benchmarking Attribute Selection Techniques for Discrete Class Data Mining,” *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 6, pp. 1437–1447, 2003.

- [8] M. B. Al Snousy, H. M. El-Deeb, K. Badran, and I. A. Al Khilil, "Suite of decision tree-based classification algorithms on cancer gene expression data," *Egypt. Informatics J.*, vol. 12, no. 2, pp. 73–82, 2011.
- [9] B. N. Lakshmi, T. S. Indumathi, and N. Ravi, "A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy," *Procedia Technol.*, vol. 24, pp. 1542–1549, 2016.
- [10] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technol.*, vol. 25, pp. 326–332, 2016.
- [11] J. S. Lee and E. S. Lee, "Exploring the Usefulness of a Decision Tree in Predicting People's Locations," *Procedia - Soc. Behav. Sci.*, vol. 140, pp. 447–451, 2014.
- [12] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," *Biomed. Signal Process. Control*, pp. 1–7, 2017.
- [13] A. Brandão, E. Pereira, F. Portela, M. F. Santos, A. Abelha, and J. Machado, "Managing Voluntary Interruption of Pregnancy Using Data Mining," *Procedia Technol.*, vol. 16, pp. 1297–1306, 2014.
- [14] H.-Y. Chen, C.-H. Chuang, Y.-J. Yang, and T.-P. Wu, "Exploring the risk factors of preterm birth using data mining," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5384–5387, 2011.
- [15] K. Paydar, S. R. Niakan Kalhori, M. Akbarian, and A. Sheikhtaheri, "A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus," *Int. J. Med. Inform.*, vol. 97, pp. 239–246, 2017.
- [16] S. Sodsee, "Predicting caesarean section by applying nearest neighbor analysis," *Procedia Comput. Sci.*, vol. 31, pp. 5–14, 2014.
- [17] I. Orozova-Bekkevold, H. Jensen, L. Stensballe, and J. Olsen, "Maternal vaccination and preterm birth: Using data mining as a screening tool," *Pharm. World Sci.*, vol. 29, no. 3, pp. 205–212, 2007.
- [18] D. Ferreira, A. Oliveira, and A. Freitas, "Applying data mining techniques to improve diagnosis in neonatal jaundice," *BMC Med. Inform. Decis. Mak.*, vol. 12, no. 143, pp. 1–6, 2012.
- [19] N. Nakayama *et al.*, "Algorithm to determine the outcome of patients with acute liver failure: A data-mining analysis using decision trees," *J. Gastroenterol.*, vol. 47, no. 6, pp. 664–677, 2012.
- [20] L. M. Taft *et al.*, "Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery," *J. Biomed. Inform.*, vol. 42, no. 2, pp. 356–364, 2009.
- [21] H. C. Koh and G. Tan, "Data mining applications in healthcare," *J. Healthc. Inf. Manag.*, vol. 19, no. 2, pp. 64–72, 2011.
- [22] G. C. Nsofor, "A Comparative Analysis Of Predictive Data-Mining," The University of Tennessee, 2006.
- [23] M. K. Saranya, R. Rathnavathy, and G. N. K. S. Babu, "A Study of Predictive Data Mining Techniques," vol. 2, no. 2, pp. 258–264, 2014.
- [24] A. Puspita and M. Wahyudi, "Algoritma C4.5 Berbasis Decision Tree untuk Prediksi Kelahiran Bayi Prematur," 2015.
- [25] Lakshmi B.N., Indumathi T.S., and N. Ravi, "A comparative study of classification algorithms for risk prediction in pregnancy," in *TENCON 2015 - 2015 IEEE Region 10 Conference*, 2015, pp. 1–6.
- [26] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [27] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree ID3 and C4.5," *Int. J. Adv. Comput. Sci. Appl.*, pp. 13–19, 2014.
- [28] A. Gupta, S. Gupta, and D. Singh, "A Systematic Review of Classification Techniques and Implementation of ID3 Decision Tree Algorithm," pp. 144–152, 2015.
- [29] R. Pandya and J. Pandya, "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning," *Int. J. Comput. Appl.*, vol. 117, no. 16, pp. 18–21, 2015.
- [30] R. Kumar and D. Verma, "Classification Algorithms for Data Mining: A Survey," *Int. J. Innov. Eng. ...*, vol. 1, no. 2, pp. 7–14, 2012.
- [31] V. Rafe and R. H. Farhoud, "A Survey on Data Mining Approaches in Medicine," vol. 4, no. 5, pp. 196–202, 2013.
- [32] T. Zhang, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods," *AI Mag.*, vol. 22, no. 2, p. 103, 2001.
- [33] G. Schwarzer, W. Vach, and M. Schumacher, "On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology," *Stat. Med.*, vol. 19, no. 4, pp. 541–561, Feb. 2000.
- [34] S. Archana and D. K. Elangovan, "Survey of Classification Techniques in Data Mining," *Int. J. Comput. Sci. Mob. Appl.*, vol. 2, no. 2, pp. 65–71, 2014.

- [35] S. Singh and P. Gupta, "Comparative study ID3, cart and C4 . 5 Decision tree algorithm: a survey," *Int. J. Adv. Inf. Sci. Technol.*, vol. 27, no. 27, pp. 97–103, 2014.
- [36] RULEQUEST RESEARCH, "C5.0: An Informal Tutorial," 2019. [Online]. Available: <https://www.rulequest.com/see5-unix.html>. [Accessed: 14-Nov-2019].
- [37] L. Auria and R. A. Moro, "Support Vector Machines (SVM) as A Technique for Solvency Analysis," *DIW Berlin Ger. Inst. Econ. Res.*, vol. No. 811, pp. 1–16, 2008.
- [38] S. Abe, *Support vector machines for pattern classification*. Manhattan: Springer, 2010.
- [39] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, Nov. 1996.